2013

Towards a Model for Predicting Intention in 3D Moving-Target Selection Tasks

Juan Sebastian Casallas
Iowa State University, casallas@iastate.edu

James H. Oliver
Iowa State University, oliver@iastate.edu

Jonathan W. Kelly
Iowa State University, jonkelly@iastate.edu

Frederic Merienne
Arts et Métiers ParisTech

Samir Garbaya
Arts et Métiers ParisTech

Towards a Model for Predicting Intention in 3D Moving-Target Selection Tasks

Juan Sebastián Casallas[1,3], James H. Oliver[1], Jonathan W. Kelly[1,2], Frédéric Merienne[3] and Samir Garbaya[3]

[1] Virtual Reality Applications Center, Iowa State University
1620 Howe Hall, Ames, Iowa, USA
[2] Department of Psychology, Iowa State University
W112 Lagomarcino Hall, Ames, Iowa, USA
[3] Institut Image, Arts et Métiers ParisTech
2 rue Thomas Dumorey, Chalon-sur-Saône, France
`{casallas,oliver,jonkelly}@iastate.edu`,
`frederic.merienne@ensam.eu,samir.garbaya@gmail.com`

Abstract. Novel interaction techniques have been developed to address the difficulties of selecting moving targets. However, similar to their static-target counterparts, these techniques may suffer from clutter and overlap, which can be addressed by predicting intended targets. Unfortunately, current predictive techniques are tailored towards static-target selection. Thus, a novel approach for predicting user intention in moving-target selection tasks using decision-trees constructed with the initial physical states of both the user and the targets is proposed. This approach is verified in a virtual reality application in which users must choose, and select between different moving targets. With two targets, this model is able to predict user choice with approximately 71% accuracy, which is significantly better than both chance and a frequentist approach.

Keywords: User intention, prediction, Fitts' Law, moving-target selection, perceived difficulty, decision trees, virtual reality

1 Introduction

Selection of moving targets is a common task in human–computer interaction (HCI) and more specifically in virtual reality (VR). Unfortunately, most of the HCI studies on selection, based on Fitts' Law [4], have focused on static targets (for a compendium, see, for example [6]). Recently, however, new performance models [1] and interaction techniques [8] have been proposed to address the specificities and difficulties of moving-target selection.

Novel moving-target selection techniques, such as *Comet* and *Ghost* [8], enhance pointing by expanding selectable targets or creating easier-to-reach proxies for each target, respectively. Nevertheless, these techniques may suffer from clutter and overlap when the number of selectable objects is increased [8]. A possible solution to these limitations, also present in static selection, is to predict

This is a manuscript of an article from *Lecture Notes in Computer Science* 8019 (2013): 13, doi: 10.1007/978-3-642-39360-0_2. Posted with permission. The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-642-39360-0_2.

the intended targets [8,15]. Unfortunately, to the authors' knowledge, current predictive techniques are tailored towards static-target selection.

Current static-target prediction techniques are based on the trajectory and velocity profiles of the pointer [13,17,21,15]. The peak accuracy rates for prediction using these techniques require a wide window of user input—at least 80% of the pointing movement—but some of them are intended to predict endpoints [13,21], rather than intended targets [17,15].

In contrast with static-target prediction techniques, this study explores the feasibility of predicting intended moving targets based only on the initial physical states of both the user and the targets, namely initial hand position, target position and target size, in a 3D selection task. To exclude factors, other than size and position, that may bias these predictions, the targets in the analyzed 3D task are kept identical in every other aspect, such as color and speed.

1.1 Identical Choices, Mental Effort and Fitts' Index of Difficulty

In the mid 90's, Christenfeld [3] conducted a series of real-life studies in which he found the middle position to be up to 75% predictive of people's choices when selecting among otherwise identical options, such as items from the same product in a supermarket or restroom stalls. In the same series of studies, he also explored route selection and found participants tended to choose based on the initial segment of the route and not on the optimal route—this was posteriorly named the *Initial Segment Strategy* [2]. Christenfeld suggested that these outcomes are consistent with the principle of minimizing *mental effort*, although he did not formalize this notion.

From a human-performance standpoint, selecting the middle choice among identical objects minimizes Fitts' Law's *Index of Difficulty* [4] (ID, see Equation 1 for its so-called Shannon Formulation [14]), since middle objects have the smallest distance (D) relative to the person and thus the smallest ID. Recent research also suggests that Fitts' ID can be related to perceived difficulty [5,12,20]. Thus, in Christenfeld's studies, people may have minimized their perceived effort by choosing the objects with the minimum ID.

$$ID = \log_2(D/W + 1) \tag{1}$$

This research explores the hypothesis that this relationship between ID and perceived effort can be used to predict user intention in 3D selection tasks. To do so, however, the influence of target distance (D) *and* width (W) on such predictions must be evaluated; as opposed to distance only, the common factor in Christenfeld's item selection studies. More importantly, it is possible that the correlation between ID and perceived effort will decrease with the addition of target motion, since the correlation between ID and selection time is reduced in moving-target selection relative to static selection [10]. Regardless, we hypothesize that ID, or another function of target size and initial distance may be predictive of user intention in moving-target selection tasks. Additionally, in accordance with the Initial Segment Strategy, we hypothesize that in the case of a sequential selection task, the first target's ID will be more predictive of user intention than the sum of IDs in the sequence.

2 Methods

2.1 Participants

Twenty-six unpaid participants, from the city of Chalon-sur-Saône, France aged 23 to 47, participated in the study. There were eighteen males and eight females; only two participants were left-handed.

2.2 Apparatus

The experiment was developed in VR JuggLua [18], a Lua wrapper for VR Juggler and OpenSceneGraph. The application was deployed in the "MoVE", a 4-surface CAVE-like virtual environment with three walls and a floor. The $3{\times}3{\times}2.67$ m environment was projected using passive, Infitec stereo [11] at $1160{\times}1050$ pixels per face. Four infrared ART cameras tracked the pose (position, P, and orientation, Q) of each participant's head and wand, using reflective markers mounted on Infitec stereo glasses and an ART Flystick2, respectively. This allowed each participant to have an adequate 3D perception and interact with the virtual world.

2.3 Procedure

Each participant was asked to stay in the middle of the MoVE ($x = 0, z = 0$) facing the front wall and was instructed to complete a series of target selection tasks. In each trial, they were presented with a horizontal array of virtual spheres of different sizes, starting in front of them and flying towards them in z. All of the spheres had the same texture, scaled accordingly to the sphere's size. Each participant was instructed to touch each sphere by extending their arms only to reach the spheres—as opposed to wait for the spheres with their arms already extended. If a sphere was touched, or if it got 0.5 m past the participant's head in z, it disappeared. Each trial ended when the participant had touched all of the spheres, or when the remaining spheres got past their head.

Visual and auditory feedback were used to indicate participant's performance. A virtual counter was placed in front of the participant at $(0, 0, -5)$, which would show the number of missed spheres during each block; the counter would be reset to zero at the beginning of each block of trials. When a participant hit a sphere, a spatialized sound would be played, co-localized with the wand position; a different spatialized sound would be played, co-localized with the overall centroid of the remaining spheres, when the spheres got past the participant's head.

At each frame of the application, the elapsed time, head pose (P_h, Q_h), wand pose (P_w, Q_w), sphere positions (P_i) and possible collisions between the wand and the spheres were recorded in a log file. The experimental setup is depicted in Fig. 1.

Fig. 1. Experimental setup with an array two spheres

2.4 Design

A within-subjects, factorial design was used, with two blocks of trials, each with a different number of conditions presented in a random order. In every trial, all of the spheres appeared 0.3 m below the participant's head and 5 m in front of them ($P_{i,y} = P_{h,y} - 0.3, P_{i,z} = -5$).

In the first block each trial had only one sphere, moving at a constant speed of 2.5 m/s in z. Factors were sphere radius ($r_1 = [0.1, 0.2]$) and sphere position (*left*: $P_{1,x} = 0.5$, *center*: $P_{1,x} = 0$, and *right*: $P_{1,x} = 0.5$). Each of the six conditions was presented to the participant in a random order until completing five trials per condition (30 total). The first block was intended only for training, so that users could become familiar with the environment and the task.

After completing the first block, the number of spheres was increased to two and velocity was decremented to 1.5 m/s in z. The spheres were positioned 0.5 m apart in x but the pair could appear offset to the *right* ($P_{1,x} = -0.5, P_{2,x} = 0$), *left* ($P_{1,x} = -0.5, P_{2,x} = 0$), or *center*ed ($P_{1,x} = -0.25, P_{2,x} = 0.25$) with respect to the user (see Fig. 2). Factors were sphere radius ($r_i = [0.1, 0.2]$) and row position (*left*, *center* and *right*). Each of the 12 conditions was presented to the participant in a random order until completing five trials per condition (60 total).

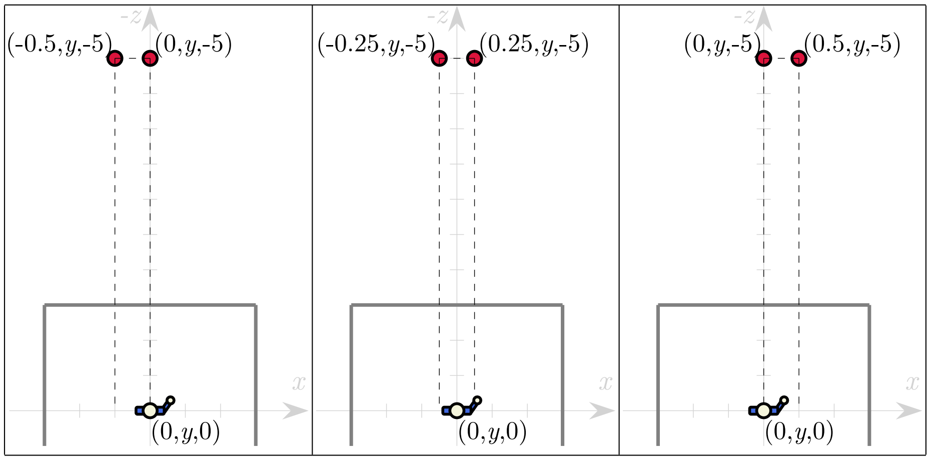

Fig. 2. Possible row positions—*left*, *center* and *right*—with respect to the user in the two-sphere block

3 Analysis

Trials in which a participant did not touch any sphere were discarded. Based on the initial wand position (P_w), sphere diameter (W_1,W_2) and initial sphere position (P_1,P_2), different values were calculated, including wand-sphere distances,

$$D_1 = |P_w - P_1| \tag{2}$$

$$D_2 = |P_w - P_2| \tag{3}$$

wand–sphere indices of difficulty,

$$ID_1 = \log_2(D_1/W_1 + 1) \tag{4}$$

$$ID_2 = \log_2(D_2/W_2 + 1) \tag{5}$$

inter–sphere distance,

$$D_{sph} = |P_2 - P_1| \tag{6}$$

inter–sphere indices of difficulty,

$$ID_{1,2} = \log_2(D_{sph}/W_2 + 1) \tag{7}$$

$$ID_{2,1} = \log_2(D_{sph}/W_1 + 1) \tag{8}$$

and total indices of difficulty

$$ID_{T1} = ID_1 + ID_{1,2} \tag{9}$$

$$ID_{T2} = ID_2 + ID_{2,1} \tag{10}$$

Using the Weka machine-learning suite [7], feature-sets $\{ID_{T1},ID_{T2}\}$, $\{ID_{1,2}, ID_{2,1}\}$, $\{ID_1,ID_2\}$ and $\{D_1,D_2,r_1,r_2\}$ were evaluated with the J48 classifier, an open source implementation of the C4.5 decision tree algorithm [19], to predict the first selected sphere.

The classifier chooses its decision nodes recursively, based on the feature that yields the greatest *Information Gain* (I)—a measure of the diminution of entropy (H, a measure of uncertainty) on the training set (S) when splitting it by the values of feature (A). In this experiment, the equations for I and H are the following:

$$I(S, A) = H(S) - H(S|A) \tag{11}$$

$$H(S) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 \tag{12}$$

$$H(S|A) = \sum_{v \epsilon Values(A)} \frac{|S_v|}{|S|} H(S_v) \tag{13}$$

where p_i is the relative frequency (see Equation 14) of sphere i (sph_i) within set S and S_v corresponds to the subset obtained by splitting S with the value v of feature A. The advantage of this classifier is that it produces easy to interpret rules, choosing the simplest decision tree from the input attributes. In this study's scope, the decision trees allowed representation and analysis of the possible participant strategies to solve each task. To avoid over-fitting to the experimental data, 10-fold cross validation was used on the generated tree models.

Finally, data were also analyzed using a frequentist approach, by calculating the relative frequency of choosing either sphere:

$$p_i = n_i/N \tag{14}$$

where n_i corresponds to the number of trials in which sph_i was chosen and N is the total number of trials. This approach allows generating a simple, one-node decision tree with an empty feature-set ($\varnothing$) that always predicts the sphere with the highest frequency.

4 Results

Participants showed an overall preference for the right sphere. The decision tree generated using the frequentist approach always predicted sph_2 as the selected sphere with approximately 64% ± 2.4% accuracy, with a 95% confidence level (see last row of Table 1).

Decision trees generated with the J48 algorithm from feature-sets 1–4 (see Table 1) yielded approximately 71% ± 2.26% accuracy on predicting the selected sphere, with a 95% confidence level, which is significantly better than both chance and a frequentist approach. Statistically, none of the tested feature-sets seemed to perform significantly better (or worse) than each other; however, the generated tree for feature-set 1 is more complex than those generated for feature-sets 2–4, making it less practical and perhaps over-fitted to the data [16] considering that the 5 non-leaf nodes were generated from only 2 attributes.

	Feature-set	Tree Size	Number of Leaves	Accuracy	95% Confidence Interval
1	ID_{T1}, ID_{T2}	9	5	70.5577%	±2.27491%
2	$ID_{1,2}$, $ID_{2,1}$	5	3	71.2062%	±2.26003%
3	ID_1, ID_2	5	3	70.9468%	±2.26605%
4	D_1, D_2, r_1, r_2	5	3	71.2062%	±2.26003%
5	$\varnothing$	1	1	63.8132%	±2.39848%

Table 1. Accuracy and 95% confidence intervals for the evaluated feature-sets

Interestingly, the fact that feature-sets 2 and 4 had the same accuracy, 95% CIs and a similar tree configuration (3 leaves out of 5 nodes) implies that they are equivalent. This may seem surprising, since the only relevant factors in the inter–sphere indices of difficulty, which compose feature-set 2, are sphere diameters (W_1, W_2)[4] (see Equations 7 and 8), whereas feature-set 4 is composed not only of sphere radii (r_1,r_2), but also of wand–sphere distances (D_1, D_2). A closer look at the generated decision tree for feature-set 4 (Fig. 3), however, shows that the decision tree included only sphere radii; wand-sphere distances (D_1, D_2) were probably ignored by the J48 algorithm on the basis of low information gain. Thus, it is safe to conjecture that the radii provide an equivalent information gain not only to feature-set 2, but also to feature-set 3, since their generated trees had similar configurations and yielded an equivalent accuracy.

The overall tendency for choosing the right sphere (sph_2) first is most likely due to the majority of the participants being right-handed; unfortunately, there weren't enough left-handed participants to evaluate the effects of handedness on the generated models. According to the decision tree generated from feature-set 4 (see Fig. 3), participants would only choose sph_1 first if sph_2 was smaller; if both spheres had the same radius, or if sph_2 was bigger, sph_2 would be selected first.

5 Discussion

Considering that decision trees were built based only on the initial position of the user's wand and the initial size and position of the spheres, predictions bore a very high accuracy. It is likely, however, that the accuracy will decrease if the number of targets is increased, but it is expected that the accuracy will still be better than chance.

Considering that the decision tree for feature-set 4 consisted only of radii, quantities from which every other feature-set is derived (see Equations 4–10)

[4] Inter-sphere distances are equal for all of the trials ($D_{sph} = 0.5$), annulling their influence on $ID_{1,2}$ and $ID_{2,1}$ and, thus, on feature-set 2

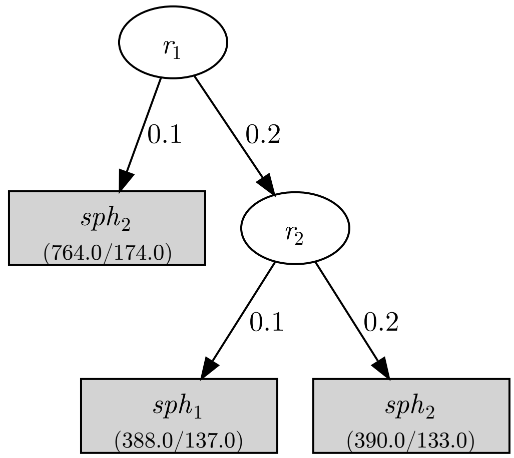

Fig. 3. Decision tree for feature-set 4, suggesting that participants based their decisions only on sphere size, with a preference for the right sphere. Leaves represent prediction outcomes (sph_1 or sph_2), while the other nodes represent tested attributes (r1 or r2). The numbers in parenthesis within the leaves represent the total number of instances that fall into that leaf, over the number of incorrectly predicted instances among these instances.

and apparently equivalent in terms of information gain to the features in other feature-sets (see the results section), suggests that size is more predictive of intended targets than every measure of Fitts' ID evaluated. This may be due to the fact that the spheres get closer to the user throughout each trial, eventually annulling the z-component of the target's distance, this corroborates Jagacinski's findings on 1D selection times on moving targets [10]. The fact that absolute horizontal sphere positions ($P_{i,x}$) did not affect user choices may suggest that users prepared their hands horizontally, while waiting for the target, or that it was more comfortable for them to reach for the right sphere first, followed by the left sphere, which seems reasonable considering that most participants were right handed.

In any case, this result suggests that participants did not have an optimal global strategy to execute their reaching tasks. Yet, this does not imply that participants optimized the initial segment either, at least as hypothesized in terms of Fitts' ID.

6 Conclusion and Future Work

The feasibility of predicting user intention in a very simple moving-target selection task was demonstrated. This approach revealed the practicality and power of using decision trees to predict user intention. Although Fitts' ID served as a good predictor of intended target selection, sphere radius seemed to yield equivalent accuracy. This suggests a very basic strategy from the users in which distance does not play an important role for choosing targets. Because the targets were moving and there was some waiting time while the target arrived, it is possible that users prepared the starting position of their wands prior to executing the pointing task.

Future work should include a greater number of spheres with different vertical positions, as well as different movement directions. Beyond size, distance and movement, this approach could be extended to consider other factors such as target semantics, if any, as well as user behaviors and gestures. The potential of using other "indices of difficulty," formulated specifically for moving-target selection [1,9,10], to predict user intention should also be explored. Finally, it should also be possible to refine decision trees in real time, to adapt the generated models to each user.

References

1. Abir Al Hajri, Sidney Fels, Gregor Miller, and Michael Ilich. Moving target selection in 2D graphical user interfaces. In Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler, editors, *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part II - INTERACT'11*, pages 141–161, Lisboa, Portugal, 2011. Springer-Verlag.
2. Jeremy N. Bailenson, Michael S. Shum, and David H. Uttal. The initial segment strategy: a heuristic for route selection. *Memory & Cognition*, 28(2):306–18, March 2000.
3. Nicholas Christenfeld. Choices from identical options. *Psychological Science*, 6(1):50–55, 1995.
4. Paul M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. (1954). *Journal of Experimental Psychology: General*, 121(3):262–269, September 1992.
5. Suzanne M. Grilli. *Perceived Difficulty in a Fitts Task*. PhD thesis, Cleveland State University, 2011.
6. Yves Guiard and Michel Beaudouin-Lafon. Fitts' law 50 years later: applications and contributions from human-computer interaction. *International Journal of Human-Computer Studies*, 61(6):747–750, December 2004.
7. Mark Hall, Hazeltine National, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA Data Mining Software : An Update. *SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.
8. Khalad Hasan, Tovi Grossman, and Pourang Irani. Comet and Target Ghost: Techniques for Selecting Moving Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '11*, pages 839–848, Vancouver, BC, Canada, 2011. ACM.
9. Errol R. Hoffmann. Capture of moving targets: a modification of Fitts' Law. *Ergonomics*, 34(2):211–220, 1991.
10. Richard J. Jagacinski, Daniel W. Repperger, Sharon L. Ward, and Martin S. Moran. A Test of Fitts' Law with Moving Targets. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 22(2):225–233, 1980.
11. H Jorke, A Simon, and M Fritz. Advanced Stereo Projection Using Interference Filters. In *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*, pages 177–180, Istanbul, Turkey, May 2008. IEEE.

12. D. Kourtis, N. Sebanz, and G. Knoblich. EEG correlates of Fitts's law during preparation for action. *Psychological research*, 76(4):514–24, July 2012.
13. Edward Lank, Yi-Chun Nikko Cheng, and Jaime Ruiz. Endpoint prediction using motion kinematics. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, pages 637–646, San Jose, CA, USA, 2007. ACM.
14. I. Scott MacKenzie. A Note on the Information-Theoretic Basis for Fitts' Law. *Journal of Motor Behavior*, 21(3):323–330, 1989.
15. Michael J. McGuffin and Ravin Balakrishnan. Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(4):388–422, 2005.
16. Tom M. Mitchell. *Machine learning*. McGraw-Hill, Boston, MA, 1997.
17. David Noy. Predicting user intentions in graphical user interfaces using implicit disambiguation. In *CHI '01 extended abstracts on Human factors in computing systems*, pages 455–456, Seattle, Washington, USA, 2001. ACM.
18. Ryan A. Pavlik and Judy M. Vance. VR JuggLua: A framework for VR applications combining Lua, OpenSceneGraph, and VR Juggler. In *2012 5th Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS)*, pages 29–35, Singapore, March 2012. IEEE.
19. John Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA, USA, 1993.
20. Andrew B. Slifkin and Suzanne M. Grilli. Aiming for the future: prospective action difficulty, prescribed difficulty, and Fitts' law. *Experimental Brain Research*, 174(4):746–53, October 2006.
21. Jonathan Wonner, Jérôme Grosjean, Antonio Capobianco, and Dominique Bechmann. SPEED : Prédiction de cibles. In *23rd French Speaking Conference on Human-Computer Interaction - IHM '11*, pages 19:1–19:4, Sophia Antipolis, France, 2011. ACM.